IES Institute of Education Sciences

APPENDIX
October 2020

Regional Educational Laboratory Southeast

At Florida State University

Are Neighborhood Factors Associated with the Quality of Early Childhood Education in North Carolina?

Appendix A. Methods

See https://go.usa.gov/xGMch for the full report.

Appendix A. Methods

The strength of any association between neighborhood characteristics and each of three outcomes (Education Standards, Program Standards, and Total Score) on the Quality Rating Improvement System (QRIS) for early childhood education sites was determined using multilevel structural equation modeling to account for the nesting of early childhood education sites within neighborhoods and to assess the direct and indirect effects of neighborhood characteristics on site quality. The study team used latent profile analysis to describe the extent to which census tracts are demographically and socioeconomically similar, using a subset of the neighborhood-level variables. The classifications derived through the latent profile analysis were used as predictors in the multilevel models to simplify interpretation of the results.

As described in the main report, the QRIS Total Score consists of a possible 7 points on the Education Standards component, 7 points on the Program Standard component, and 1 Quality Point. Sites can earn the Quality Point in several ways, related to either Education Standards (such as having all lead teachers and at least 75 percent of all teachers with an associate's degree or higher and having all teachers or family child care site providers complete 20 additional hours of annual training) or Program Standards (such as reducing group size, reducing staff–child ratio, and reducing the licensure capacity of the number of preschool children).

Primary and supplemental sample selection procedures

At the time the study team received the dataset from the North Carolina Division of Child Development and Early Education (DCDEE) in December 2017, 6,271 sites were licensed to provide child care or early childhood education in the state. That dataset matched the full set of licensed sites and the site-level data publicly available on the DCDEE website at the time; the data were provided as a file, thereby avoiding unnecessary data entry by the study team. Of the 6,271 sites, 4,531 (72 percent) were "centers" or "other" free-standing child care centers and sites within or administered by schools or religious institutions such as churches, and 1,740 (28 percent) were family child care homes.

Several eligibility criteria were applied to derive the analytical sample. First, any site whose geolocation could not be determined was excluded. (This process is described in the next section.) Second, sites had to meet the age range criterion of serving at least one age within ages 3–5 for children not enrolled in kindergarten or exclusively in a before- or after-school program. The age range for licensed child care sites spans birth to age 12; sites were retained if this range included age 3 or 4 or both; no otherwise eligible sites served only 5-year-old children. The age range variable within the dataset was used to make the majority of these determinations. Some early childhood education site websites also were used to eliminate programs that provided exclusively before- or after-

school care. Collectively, these criteria excluded 520 sites, yielding a retained sample of 5,735 (about 92 percent of the full set): 4,019 “center” or “other” sites and 1,716 family child care home sites. Finally, 481 sites were excluded because the data did not include all three quality rating scores used as outcome measures. Of these, 339 sites coded as “other” were excluded because they were religion-sponsored sites that were exempt from requirements to participate in the Star Rated License process, and 142 were sites with temporary or provisional licenses that had not yet received their ratings. Thus, the final analytic sample included 5,254 sites. Geolocation procedures were applied to the 481 sites that did not meet this third eligibility criterion for inclusion in the analytic sample because the sites were included in the procedures for the counts of early childhood education sites, which is described below.

In addition to the sites included in the analytic sample, the study team sought to identify additional sites in North Carolina that were neither licensed by the state nor included on the original list of 6,271 sites. Nonduplicative sites were identified from lists of public schools with Title I prekindergarten classrooms; Head Start and Early Head Start sites serving children ages 3–5; and publicly available lists of private schools that had prekindergarten classrooms—for example, schools accredited by the Southern Association of Independent Schools. These North Carolina early childhood education sites were not active participants in the quality rating system at the time the study team acquired the original master list. The study team identified 345 sites after manually checking for duplicates with the original master list and analytic sample. Of these, 324 passed the geolocation and more rigorous geocoding-based processes for removing duplicates. These sites were included in the density counting. Thus, the collective number of sites included for calculating counts of early childhood education sites within each census tract and within all tracts adjacent to each target tract was 6,059.

Identifying eligible postsecondary institutions

An initial set of 930 postsecondary institutions in North Carolina and the adjacent states of Georgia, South Carolina, Tennessee, Virginia, and West Virginia was obtained from the Integrated Postsecondary Education Data System maintained by the National Center for Education Statistics (n.d.). All institutions on this list were subject to eligibility screening.

A geographic information system was used to select all the institutions in North Carolina or within 50 miles of the state border. The U.S. Census Bureau and transportation experts consider a distance of 50 miles or longer to be an extra-long commute (Rapinoe & Fields, 2013), so sites further than 50 miles from the border were excluded. For the remaining sites the study team searched each institution’s website to determine relevance. Institutions with a singular career orientation, such as cosmetology and culinary institutions, were excluded immediately, along with a small number of institutions that were closed.

For the retained sites each website was searched to identify whether the institution offered an eligible major at the associate’s, bachelor’s, or graduate level or nondegree certification in early childhood or child care instruction. Searches of program and degree guides, as well as custom searches, were used to maximize the likelihood of identifying all relevant institutions. Eligible majors included early childhood education, elementary education, special education, and degrees with similar titles. Institutions also were retained if the education track was a minor but led to eligibility for an age-appropriate teaching certification or credential in the state. Approximately 25 percent of all site institutions were randomly selected for review by two coders to ascertain inter-rater reliability, which was greater than 90 percent. The lead author reviewed the coding for all institutions and resolved any institutions for which there was a discrepancy or uncertainty. Of the original 930 potential institutions, the study team retained 199 eligible institutions. The geocoding process described in the next section led the team to hand check 22 institutional addresses, all of which were resolved successfully.

Geocoding procedures

The master list of early childhood education sites from DCDEE's quality rating dataset, the list of supplemental early childhood education sites, and the list of postsecondary institutions all contained a street address for each facility. The same geocoding process described here was used for each dataset.

Geocoding converts a text address into latitude and longitude coordinates. Modern geocoders contain vast catalogues of known addresses with corresponding latitude and longitude points. The goal is to match a particular site's address to an address in the geocoder. If the site's address is not already in the geocoder, an approximate latitude and longitude point can be found by interpolating a location between known points. Geocoders vary in the number of already known locations, their ability to interpret the text of the address, and their interpolation approach. Because geocoders have strengths and weaknesses, results were retrieved from both Google and Bing geocoders, after which a decision tree was used to determine the final coordinates for each address.

Expected match rates for geocoding depend on many factors, the most important being the quality of the address text, the ability of the geocoder to interpret that text, and the number of addresses the geocoder knows. The completer and more accurate the address text, the more likely that the geocoder will match the address to a location (Cui, 2013). It is reasonable to expect that a list of commercial addresses will be of high quality because business proprietors have an incentive to make it easy for potential customers to find them (Folch et al., 2018). Early childhood education homes and centers may not have as many drop-in customers as restaurants do, but they still have an incentive to be correctly located for parents looking for early childhood education sites.

The decision tree classified each result as either low confidence or high confidence based on the accuracy and match quality. "Rooftop" accuracy means that the coordinates are at a known address, and "interpolated" means that the coordinates are estimated based on nearby addresses. Match quality describes how well the geocoders interpreted the input address text. Results with a low-quality or partial match from either Google or Bing were immediately passed to the low confidence group. Results that were considered high confidence had three features: they had rooftop accuracy from either Google or Bing, they had at least interpolated accuracy from the other geocoder, and the two geocodes were less than 200 meters apart. The final coordinates came from the rooftop geocode; if both geocoders returned rooftop accuracy, then Bing was chosen. Addresses that passed the match quality test but not the accuracy test were considered low confidence.

Based on each geocoder's definitions of accuracy, 92 percent of North Carolina early childhood education sites could be found with rooftop accuracy by Google's definitions and 90 percent by Bing's definitions. From this baseline, the study team's cross-validation approach has two advantages. First, it leverages the geocoders' different sets of known addresses. In a Venn diagram sense the two could have different strengths in their known addresses and thus combined could match more addresses than either could individually. Second, the study team validated one geocoder against the other by including the interpolations, which necessarily constrains the number of valid matches. Thus, the study team used both an expansive and a robust method for matching. With this approach, the total automated match rate for the entire state was 89 percent. This rate compares favorably to other geocoding studies: a study that sampled six North Carolina counties had match rates ranging from 47 percent (Sampson County) to 92 percent (Durham County) for place-of-death addresses (Edwards et al., 2014).

Additional steps were taken to increase this already high match rate. Once the study team identified low confidence sites, it manually reviewed the sites to determine which coordinates to use. Research assistants were trained to inspect Google StreetView and Google satellite imagery for indicators of an early childhood education site, such as playground equipment, signage, or school-related vehicles. The small set of sites that were not confirmed by this approach were contacted by phone to gather more information, such as a corrected address or local landmarks, that could help with finding it. Most sites, although not all, provided updated information that sufficiently clarified the address. If more information was not found, the site was eliminated from the final dataset.

As described above, the additional eligibility criteria led the study team to retain a very high percentage of sites (that is, less than 1 percent were excluded based solely on inconclusive location).

Calculating counts and distance

Counts of the number of early childhood education sites from the master list in each census tract were gathered by spatially linking the census tract identifier to each early childhood education site in that tract. The sites were then grouped by the tract identifier, creating a count of sites in each tract. The same process was used to produce counts of the additional, nonstudy sites in each tract. The total count used in the analyses was the sum of the study and nonstudy counts. The study team also summed the number of sites in the tracts neighboring each tract. The adjacent count for a tract is the sum of study and nonstudy sites across neighboring tracts.

The list of postsecondary education institutions was obtained from the Integrated Postsecondary Education Data System (National Center for Education Statistics, n.d.), and each institution was reviewed for eligibility and geocoding validation as described above. Using the longitude and latitude coordinates for the postsecondary institutions and those for the early childhood education sites included in the study, the study team determined the nearest postsecondary institution for each early childhood education site in the study. This distance in meters was used in subsequent analyses.

Additional details on the analyses

Latent profile analysis using a subset of the neighborhood level variables was employed to describe the extent to which census tracts were demographically and socioeconomically similar. The classifications derived through the latent profile analysis were used as predictors in the subsequent multilevel models to simplify interpretation of results. This approach was supported by sensitivity analyses (described later).

Latent profile analysis. Prior to estimating the multilevel models, the study team used latent profile analysis to combine the large set of demographic and socioeconomic status variables from the census tract into a parsimonious set of dummy variables for use in the regression analyses. This analysis included all populated census tracts in the state to ensure statewide representation.

Latent profile analysis is typically used to classify individuals into groups based on, for instance, their responses on a single exam or their scores on multiple exams. Although this type of analysis has been used predominantly for diagnostic purposes in psychology and marketing, it is an emerging descriptive classification technique in education (Koon et al., 2014). In this case the following variables were used to group census tracts that were demographically and socioeconomically similar:

- Proportion of the population that is African American, not Hispanic.
- Proportion of the population that is Asian, not Hispanic.
- Proportion of the population that is other race/ethnicity, not Hispanic.
- Proportion of the population that is Hispanic
- Proportion of the population that is below the federal poverty line.
- Proportion of the population older than age 16 that was unemployed for the past 12 months.
- Proportion of the population older than age 25 that does not have a high school diploma.
- Proportion of the population older than age 25 that has a high school diploma as highest education.
- Median household income.

- Proportion of households receiving public assistance income or food stamps/Supplemental Nutrition Assistance Program benefits.

The following variables were used in the summary of the profiles but not in the latent profile analysis due to multicollinearity: proportion of population that is White, not Hispanic; proportion of population that is older than age 5 speaking only English; and the Gini inequality index.

A basic multivariate latent profile analysis model can be represented by the following equation (Pastor et al., 2007):

$$f(y_i|\theta) = \sum_{k=1}^{K} \pi_k f_k(y_i|u_k \Sigma_k),$$

where y_i is the multivariate distribution of cluster indicators (census tract variables) for census tract i (with the number of clusters represented by k), θ is the unique set of model parameters to be estimated within each cluster, and π_k is the weight given to each cluster. The weights are constrained to be non-negative and to sum to 1. Each cluster distribution is defined by u_k (the mean vector) and Σ_k (the covariance matrix).

Multiple indices reported by the Mplus program (Muthén & Muthén, 2017) were used to determine the most appropriate number of profiles for the data (table A1). The indices include Akaike information criteria and Bayesian information criteria (Kaplan, 2000), with smaller values preferred. Also, entropy was used on a scale of 0 to 1 with higher values preferred (Ramaswamy et al., 1993). Finally, the Lo-Mendell-Rubin likelihood ratio test (Lo et al., 2001) and a parametric bootstrapped likelihood ratio test (McLachlan & Peel, 2000) were used, with a significant value on these tests indicating that the model with *K* classes is preferred to the model with *K*–1 classes (which is nested within the former model).

Table A1. Summary of model fit indices for latent profile analysis

Number of classes	Akaike information criteria	Bayesian information criteria	Adjusted Bayesian information criteria	Entropy	Lo-Mendell-Rubin likelihood ratio test (*p*-value)	Bootstrapped likelihood ratio test (*p*-value)
1	22,955.25	22,841.60	22,905.15	1.00	na	na
2	28,077.65	27,901.49	27,999.98	0.84	0.20	0.00
3	31,630.65	31,391.98	31,525.42	0.91	0.00	0.00
4	34,158.84	33,857.67	34,026.06	0.93	0.60	0.00
5	35,674.65	35,310.97	35,514.31	0.91	0.22	0.00

na is not applicable.
Note: All models converged successfully with the best log likelihood value replicated with start values of 200 and 40 and 400 and 80.
Source: Authors' analysis using data from the American Community Survey.

A three-class model was selected as the best fit to the data. Moving to four classes resulted in a nonsignificant *p*-value for the Lo-Mendell-Rubin likelihood ratio test. In addition to fitting the data well, the three-class model resulted in a solution that lent itself well to interpretation.

Individual posterior probabilities were saved for each class and used to put each census tract in a single class. The probability of each census tract's class membership can be averaged by class to further evaluate the model results. The average latent class probabilities for most likely latent class membership ranged from .95 to .96, suggesting good classification accuracy (table A2).

Table A2. Average latent class probabilities for most likely latent class membership, by latent class

Latent class membership	Latent class 1 probability	Latent class 2 probability	Latent class 3 probability
1	.96	.00	.05
2	.00	.95	.05
3	.02	.02	.96

Source: Authors' analysis using data from the American Community Survey.

Multilevel structural equation modeling. A multilevel structural equation modeling (MSEM) framework was used to test the extent to which neighborhood characteristics were associated with the quality of early childhood education sites in North Carolina (research question 1). The hierarchical nature of these models accounts for the way in which early childhood education sites are nested within neighborhoods as well as the assessment of the direct effect of neighborhood characteristics on site quality. The MSEM approach is an improvement over standard multilevel models because it separates the between-cluster and within-cluster effects instead of combining them into single coefficients (Preacher et al., 2010). All multilevel models were analyzed in Mplus (Muthén & Muthén, 2017). Preliminary analyses were conducted in SPSS to inform the multilevel models.

A model-building approach was used to sequentially examine models and create the most parsimonious model for each outcome. In the model-building process the study team first focused on establishing the level 1 model, or the site-level model. If the level 1 model is mis-specified, incorrect parameter estimates and errors of inferences can occur in the upper levels of the model (McCoach & Kaniskan, 2010). The order of variable entry was determined through stepwise regression in exploratory data analyses, using the default settings in SPSS (that is, a variable is entered into the model if the significance level of its *F*-value is less than .05 and is removed if the significance level is greater than .10). All available variables at the level of consideration were entered into the MSEM model in the same order as identified separately through the stepwise regression analyses, except the variables specific to research questions 2 and 3 (the count variables at level 2 and the distance to the nearest postsecondary institution at level 1). The variables added to each model following the order of variable entry identified through stepwise regression were also the variables that had the largest correlations with the respective quality rating score. The exception was the low correlation between the variable indicating whether the site participates in the school readiness subsidy program and the Program Standards score. However, the magnitude of this correlation increased after facility type was controlled for, making the addition of this variable to the model consistent with the order of variable entry identified through stepwise regression.

After the level 1 model was decided in Mplus, the census tract and county variables were added at level 2 (also in Mplus). The MSEM models that were considered in this process are listed in table A3, along with information on the model fit indices and variance explained at each level of the model.

For the Education Standards outcome, three level 1 predictors provided the best fit: minimum age served, participation in the subsidy program, and maximum age served. The addition of national accreditation did not substantially improve the overall model fit, and no additional level 1 predictors were tested. Next, all of the level 2 variables were considered as candidates for entry at level 2, including the profile of the census tract derived through the latent profile analysis. For the Education Standards outcome the order of entry began with average elementary school performance on statewide accountability measures at the county level. The addition of this variable did not improve the overall model fit. No additional variables were added to the model.

Table A3. Multilevel structural equation models considered in the model-building process, by quality rating score

Model	Intraclass correlation coefficient	Akaike information criteria (AIC)	Bayesian information criteria (BIC)	Sample-size adjusted BIC	Level 1 variance	Level 2 variance
Education Standards						
Null	.042	22,045.89	22,065.59	22,056.06	3.74	0.16
Model 1: Null model plus minimum age served	.045	21,691.02	21,717.29	21,704.58	3.47	0.17
Model 2: Model 1 plus participation in subsidized care	.030	20,877.79	20,910.62	20,894.74	3.00	0.11
Model 3: Model 2 plus maximum age served ***(research question 1)***	**.036**	**20,599.96**	**20,639.36**	**20,620.29**	**2.82**	**0.14**
Model 4: Model 3 plus national accreditation	.036	20,572.96	20,618.92	20,596.68	2.80	0.14
Model 5: Model 3 plus average county school performance	.035	20,589.69	20,635.66	20,613.42	2.82	0.13
Model 6: Model 3 plus density variables ***(research question 2)***	.035	20,598.20	20,650.73	20,625.31	2.82	0.13
Model 7: Model 3 plus postsecondary distance ***(research question 3)***	.036	20,601.39	20,647.36	20,625.11	2.82	0.14
Program Standards						
Null	.084	23,780.36	23,800.06	23,790.53	5.00	0.46
Model 1: Null model plus center type	.055	22,494.71	22,520.97	22,508.26	3.96	0.30
Model 2: Model 1 plus maximum age served	.056	21,943.45	21,976.29	21,960.40	3.54	0.30
Model 3: Model 2 plus participation in subsidized care	.050	21,759.74	21,799.14	21,780.07	3.43	0.27
Model 4: Model 3 plus minimum age served	.050	21,664.87	21,710.84	21,688.59	3.37	0.27
Model 5: Model 4 plus capacity of the first shift ***(research question 1)***	**.047**	**21,580.45**	**21,632.98**	**21,607.56**	**3.32**	**0.26**
Model 6: Model 5 plus national accreditation	.047	21,570.26	21,629.36	21,600.76	3.31	0.26
Model 7: Model 5 plus census tract total population	.046	21,566.00	21,625.10	21,596.50	3.33	0.24
Model 8: Model 5 plus density variables ***(research question 2)***	.046	21,561.75	21,627.42	21,595.64	3.32	0.24
Model 9: Model 5 plus postsecondary distance ***(research question 3)***	.047	21,582.32	21,641.42	21,612.82	3.32	0.26
Total Score						
Null	.051	28,977.54	28,997.24	28,987.70	13.87	0.72
Model 1: Null model plus center type	.038	27,978.81	28,005.07	27,992.36	11.51	0.55
Model 2: Model 1 plus maximum age served	.043	27,514.69	27,547.52	27,531.63	10.43	0.62
Model 3: Model 2 plus participation in subsidized care	.029	26,813.37	26,852.77	26,833.70	9.24	0.41
Model 4: Model 3 plus minimum age served	.028	26,599.98	26,645.95	26,623.70	8.86	0.40
Model 5: Model 4 plus capacity of the first shift ***(research question 1)***	**.028**	**26,526.97**	**26,579.50**	**26,554.08**	**8.73**	**0.40**
Model 6: Model 5 plus home pick-up services	.028	26,500.09	26,559.19	26,530.60	8.68	0.40
Model 7: Model 5 plus census tract total population	.027	26,517.29	26,576.39	26,547.79	8.74	0.37
Model 8: Model 5 plus density variables ***(research question 2)***	.028	26,529.59	26,595.26	26,563.48	8.73	0.40
Model 9: Model 5 plus postsecondary distance ***(research question 3)***	.028	26,528.63	26,587.73	26,559.13	8.73	0.40

na is not applicable.
Note: Models in bold represent the final model for research questions 1, 2, and 3.
Source: Authors' analyses using data from the North Carolina Division of Child Development and Early Education, the American Community Survey 2012–16, and other publicly available sources.

The model-building process for the Program Standards quality rating score and Total Score followed the same procedures as those for the Education Standards quality rating score. The final model for both of these outcomes included five level 1 predictors: center type, maximum age served, participation in the subsidy program, minimum age served, and capacity of first shift. After site-level variables were controlled for, no additional variables at level 1 or 2 were found to improve model fit, including the profile of the census tract. The fixed and random effects values for each final model are in table A4. Because no direct effects of neighborhood-level variables were included in the final models for any of the three outcomes, more complex models exploring moderation or mediation of neighborhood-level variables by site-level variables were not pursued.

The final model for each outcome for research question 1 was then used as the base model for research question 2. The density covariates (the count variables) were added as neighborhood-level covariates to test the existence of a unique effect on each quality rating score. The model fit did not improve with these additional variables, meaning that density does not have an important effect on the quality outcomes (see table A3).

Similar to the case for research question 2, the final model resulting from research question 1 was used as the base model for the same dependent variable in research question 3. The covariate representing distance to the nearest eligible postsecondary institution was added as a site-level covariate to test the existence of a unique effect on each quality rating score. The model fit did not improve with this additional variable, meaning that distance to postsecondary institution does not have a significant effect on the quality outcomes (see table A3).

Table A4. Fixed and random effects for final models under research question 1

	Fixed effects			Random effects			
Variable	Estimate[a]	Standard error	*p*-value[b]		Name	Variance component	*p*-value[b]
Education Standards (model 3)							
(Intercept)	5.55	0.17		Census tract	(Intercept)	0.14	.00
Minimum age served	0.40	0.03	.00	Residual		2.82	.00
Participation in subsidized care	2.16	0.09	.00				
Maximum age served	–0.21	0.02	.00				
Program Standards (model 5)							
(Intercept)	5.15	0.22	.00	Census tract	(Intercept)	0.26	.00
Center-based site	1.12	0.09	.00	Residual		3.32	.00
Maximum age served	–0.28	0.02	.00				
Participation in subsidized care	1.11	0.08	.00				
Minimum age served	0.33	0.04	.00				
Capacity of first shift	0.01	0.00	.00				
Total Score (model 5)							
(Intercept)	11.51	0.38	.00	Census tract	(Intercept)	0.40	0.00
Center-based site	1.10	0.13	.00	Residual		8.73	0.00
Maximum age served	–0.51	0.03	.00				
Participation in subsidized care	3.51	0.16	.00				
Minimum age served	0.75	0.07	.00				
Capacity of first shift	0.01	0.00	.00				

a. Unstandardized.

b. Significant at the .001 level.

Source: Authors' analyses using data from the North Carolina Division of Child Development and Early Education, the American Community Survey 2012–16, and other publicly available sources.

Additional detail on site-level predictors included in MSEM analyses. As described in the main report and as shown in tables A3 and A4, a subset of the site-level characteristics investigated were significant predictors of Education Standards, Program Standards, or Total Score or all three. The site-level variables that consistently predicted differences across sites included age range of children served by the site and whether the site participated in the school readiness subsidy program for children whose families met eligibility criteria, which are related primarily

to family income (North Carolina Division of Child Development and Early Education, n.d.). The associations between the two child age variables and the three quality rating scores suggest that sites serving a narrower age range generally have higher quality.

As the minimum age served increases by one year, quality rating scores rise 0.40 point for the Education Standards score, 0.33 point for the Program Standards score, and 0.75 point for the Total Score (see table A4). As the maximum age served increases by one year, however, quality rating scores decline 0.21 point for the Education Standards score, 0.28 point for the Program Standards score, and 0.51 point for the Total Score. The association between participation in the school readiness subsidy program and the three quality rating scores was positive, with an increase of 2.16 points for the Education Standards score, 1.11 points for the Program Standards score, and 3.51 points for the Total Score (see table A4).

Sensitivity analyses. The best-fitting latent profile analysis model produced just three profile groups of neighborhoods/census tracts, and most of the census tracts were included in profile 3. When considered in the MSEM models, profile membership did not explain variability in the quality rating scores over and above the variability explained by the already included key site-level variables. Additional MSEM models were analyzed to test the contribution of individual tract-level variables in lieu of considering them as represented by profile membership. All census tract variables were added to the best-fitting level 1 model for each of the three quality rating scores. For all three scores the intraclass correlation coefficient did not change from the best-fitting level 1 model, and the models being tested increased the Bayesian information criteria by 42–47 points. Given these results, census tract-level variables were retained within profile membership instead of being considered separately.

The study team used stepwise regression and forward selection procedures in exploratory data analyses to guide the order in which variables were entered into the MSEM models. In addition, the study team consulted correlation tables to evaluate the consistency of the results of the variable selection procedures and the magnitude of the correlations between the variables and the outcomes. To ensure that the goal of creating the most parsimonious model for each outcome was not overly restrictive, the study team estimated MSEM models with all of the potential level 1 variables (full model) and compared these model results with the models guided by stepwise regression and restricted by model fit. For Education Standards, the full model explained 1.6 percent of the level 1 variance over that of the restricted model and resulted in a decrease in the intraclass correlation coefficient by 0.004, to 0.032. The study team found smaller differences in the full and restricted models for the Program Standards and Total Scores outcomes, with 0.3 percent additional level 1 variance explained by the Program Standards full model and 0.8 percent of additional variance explained by the Total Scores full model. For these two outcomes the intraclass correlation coefficients did not differ between models. These sensitivity analyses support the specification of the final level 1 models used in the study, as the addition of all level 1 variables explains less than 2 percent of additional variation in the outcomes while increasing the complexity of the models.

References

Cui, Y. (2013). A systematic approach to evaluate and validate the spatial accuracy of farmers market locations using multi-geocoding services. *Applied Geography*, *41*(1), 87–95.

Edwards, S. E., Strauss, B., & Miranda, M. L. (2014). Geocoding large population-level administrative datasets at highly resolved spatial scales. *Transactions in GIS*, *18*(4), 586–603.

Folch, D. C., Spielman, S. E., & Manduca, R. (2018). Fast food data: Where user-generated content works and where it does not. *Geographical Analysis*, *50(2)*, 125–140.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Sage Publications.

Koon, S., Petscher, Y., & Foorman, B. R. (2014). *Beating the odds: Finding schools exceeding achievement expectations with high-risk students* (REL 2014–032). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778.

McCoach, D. B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in Psychology, 1*(17), 1–12.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons, Inc.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th edition). Muthén & Muthén.

National Center for Education Statistics. (n.d.). Integrated Postsecondary Education Data System. Retrieved March 9, 2018, from https://nces.ed.gov/ipeds/use-the-data.

North Carolina Division of Child Development and Early Education. (n.d.). How Does Child Care Financial Assistance in North Carolina Work? Retrieved April 8, 2019, from https://ncchildcare.ncdhhs.gov/Services/Financial-Assistance.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8–47. https://eric.ed.gov/?id=EJ751860.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://eric.ed.gov/?id=EJ897279.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, *12*(1), 103–124.

Rapinoe, M. A., & Fields, A. K. (2013, November). *Mega commuters in the U.S.: Time and distance in defining the long commute using the American Community Survey.* Paper presented at the Association for Public Policy Analysis and Management Fall Conference, Washington, DC.